# The Kiki-Bouba Challenge

*Algorithmic composition for content-based MIR Research & Development*

Sturm, Bob L.; Collins, Nick

*Published in:*
Proceedings of the 15th International Society for Music Information Retrieval Conference, ISMIR 2014

*Publication date:*
2014

*Document Version*
Early version, also known as pre-print

Link to publication from Aalborg University

*Citation for published version (APA):*
Sturm, B. L., & Collins, N. (2014). The Kiki-Bouba Challenge: Algorithmic composition for content-based MIR Research & Development. In *Proceedings of the 15th International Society for Music Information Retrieval Conference, ISMIR 2014* (pp. 21-26). dblp Computer Science Bibliography. http://composerprogrammer.com/kikibouba.html

# THE KIKI-BOUBA CHALLENGE: ALGORITHMIC COMPOSITION FOR CONTENT-BASED MIR RESEARCH & DEVELOPMENT

**Bob L. Sturm**
Audio Analysis Lab, Aalborg University, Denmark
bst@create.aau.dk

**Nick Collins**
Dept. Music, Durham University, UK
nick.collins@durham.ac.uk

## ABSTRACT

We propose the "Kiki-Bouba Challenge" (KBC) for the research and development of content-based music information retrieval (MIR) systems. This challenge is unencumbered by several problems typically encountered in MIR research: insufficient data, restrictive copyrights, imperfect ground truth, a lack of specific criteria for classes (e.g., genre), a lack of explicit problem definition, and irreproducibility. KBC provides a limitless amount of free data, a perfect ground truth, and well-specifiable and meaningful characteristics defining each class. These ideal conditions are made possible by open source algorithmic composition — a hitherto under-exploited resource for MIR.

## 1. INTRODUCTION

Before attempting to solve a complex problem, one should approach it by first demonstrably solving simpler, well-defined, and more restricted forms, and *only then* increase the complexity. However, there are key problems of research in content-based music information retrieval (MIR) [8] where this has yet to be done. For example, much of the enormous amount of research that attempts to address the problem of music genre recognition (MGR) [26] has started with genre in the "real world" [30]. The same is seen for research in music mood recognition [28, 29, 37], and music autotagging [6]. On top of this, the problem of describing music using genre, mood, or tags in general, has rarely, if ever, been explicitly defined [32].

In lieu of an explicit definition of the problem, the most common approach in much of this research is to implicitly define it via datasets of real music paired with "ground truth." The problem then becomes reproducing as much of the "ground truth" as possible by pairing feature extraction and machine learning algorithms, and comparing the resulting numbers to those of other systems (including humans). Thousands of numerical results and publications have so far been produced, but it now appears as if most of it has tenuous relevance for *content-based* MIR [3, 27, 30, 31, 34]. The crux of the argument is that the lack of scientific validity in evaluation in much of this work [3, 27, 30] has led to the development of many MIR systems that appear as if they are "listening" to the music when they are actually just exploiting confounded characteristics in a test dataset [31]. Thus, in order to develop MIR systems that address the goal of "making music, or information about music, easier to find" [8] in the real-world, there is a need to first demonstrably solve simple, well-defined and restricted problems.

Toward this end, this paper presents the "Kiki-Bouba Challenge" (KBC), which is essentially a simplification of the problem of MGR. On a higher level, we propose KBC to refocus the goals in content-based MIR. We devise KBC such that solving it is unencumbered by six significant problems facing content-based MIR research and development: 1) the lack of formal definition of retrieving information in recorded music; 2) the large amount of data necessary to ensure representativeness and generalization for machine learning; 3) the problem of obtaining "ground truth"; 4) the stifling affect of intellectual property (e.g., music copyright) on collecting and sharing recorded music; 5) the lack of validity of standard evaluation approaches of systems; and 6) a lack of reproducible research. KBC employs algorithmic composition to generate a limitless amount of music from two categories, named *Kiki* and *Bouba*. Music from each category are thereby free from copyright, are based in well-defined programs, and have a perfect ground truth. Solving KBC represents a veritable contribution of content-based MIR research and development, and promises avenues for solving parallel problems in less restricted and real-world domains.

Instead of being merely the reproduction of a "ground truth" of some dataset, the MIR "flagship application" of MGR [4] — and that which KBC simplifies — has as its principal goals the *imitation of the human ability to organize, recognize, distinguish between, and imitate genres used by music* [28]. To "imitate the human ability" is not necessarily to replicate the physiological processes humans use to hear, process and describe a piece of music, but merely to describe as humans do a piece of music *according to its content*, e.g., using such musically meaningful attributes as rhythm, instrumentation, harmonic progression, or formal structure. Solving the problem of MGR means creating an artificial system that can work with music like humans, but unencumbered by human limitations.

The concept of genre [12, 13, 16] is notoriously difficult to define such that it can be addressed by algorithms [23]. Researchers building MGR systems have by and large posed the problem, implicitly or explicitly, from an Aristotelean viewpoint, i.e., "genre" is a categorization of music just as "species" is a categorization of living things, e.g., [5,33]. [1] The problem then is to automatically learn the characteristics that place a piece of music on one branch of a taxonomy, distinguish it from a piece of music on a different branch, and avoid contradiction in the process [7]. Researchers have combined signal processing and machine learning with datasets of real music recordings in hopes that the resulting system can discover Aristotelean criteria by which music can be categorized according to genre. The majority of the resulting work, however, documents how much "ground truth" an algorithm replicates in benchmark datasets [26], but rarely illuminates the criteria a system has learned and is using to categorize music [27]. The former quantity is meaningless when the latter is senseless.

In the next section, we discuss the use of algorithmic music composition for data generation. Then we present KBC in its most general form. We follow this with a concrete and specific realisation of KBC, available at the relevant webpage: `http://composerprogrammer.com/kikibouba.html`. We present an unacceptable solution to KBC, and discuss aspects of an acceptable solution. We conclude this paper with a discussion of KBC, and how it relates to content-based MIR in the "real world."

## 2. ALGORITHMIC MUSIC COMPOSITION FOR GENERATING DATA

Algorithmic composition [1,9,19,21,22,25,36] has a long history back to mainframe computer experiments in the mid 1950s, predating by a decade MIR's first explicit paper [17]. Ames and Domino [2] differentiate *empirical style modeling* (of historic musical styles) and *active style synthesis* (of novel musical style). In the practical work of this article we concentrate more on the latter, but there is a rich set of techniques for basing generation of music on models trained on existing musical data. Many musical models deployed to capture regularities in data sets are generative, in that a model trained from a corpus can generalise to production of new examples in that style [11].

Though anticipated by some authors, it is surprising how few studies in computer music have utilised algorithmic composition to create the ground truth. Although [24] present a four category taxonomy of algorithmic composition, they do not explicitly discuss the option of using algorithmic composition to produce data sets. The closest category is where "theories of a musical style are implemented as computer programs" [24], essentially empirical style modeling as above.

Sample CD data, especially meta-data on splices, have also rarely been used. But the advantage of algorithmic composition techniques are the sheer volume of data which can potentially be generated, and appropriately handled should be free of the copyright issues that plague databases of music recordings and hinder research access.

We believe that algorithmic generation of datasets within a framework of open source software has the following potential benefits to MIR and computer music analysis:

- Limitless data set generation, with perfect ground truth (the originating program is fully accessible, and can be devised to log all necessary elements of the ground truth during generation. Random seeds can be used to recover program runs exactly as necessary)
- A fully controlled musical working space, where all assumptions and representational decisions are clear
- Copyright free as long as license free samples or pure synthesis methods are utilised, under appropriate software licensing
- Established data sets can be distributed free of the originating software once accepted by the community, though their origins remain open to investigation by any interested researcher

The greatest issue with dependence on algorithmic generation of music is the ecological validity of the music being generated. A skeptic may question the provenance of the music, especially with respect to the established cultural and economically proven quality of existing human driven recorded music production. Nonetheless, humans are intimately involved in devising algorithmic composition programs. We believe that there is place for expert judgement here, where experts in algorithmic composition can become involved in the process of MIR evaluation. The present paper serves as one humble example; but ultimately, a saving grace of any such position is that the generation code is fully available, and thus accessible to reproduction and evaluation by others.

## 3. THE KIKI-BOUBA CHALLENGE

We now present KBC in its most general form: *develop a system that can organize, recognize, distinguish between, and imitate Aristotelean categories of "music."* We define these in the subsections below, after we specify the domain.

### 3.1 Domain

The music universe of KBC is populated by "music" belonging to either one of two categories, *Kiki* and *Bouba*. [2] In KBC, music from either category is algorithmically composed such that there is available a limitless number of recordings of music from both categories, and which are entirely unencumbered by copyrights. A music recording from this universe therefore embeds music from *Kiki* and not from *Bouba*, or vice versa, for several reasons that are neither ambiguous nor disputable, and which can be completely garnered from the music recording. The ground truth of a dataset of recordings of music from the music universe then is absolute. Note that a music recording need not be an audio recording, but can be a notated score, or other kind of representation. Now, given that this is ideal

[1] This of course belies the profound issues that biologists face in recognizing "speciation" events [10].

[2] Shapes named "Kiki" and "Bouba" (the two are spiky and rounded, respectively) were originally introduced in gestalt psychology to investigate cross-cultural associations of visual form and language [18, 20]. Our example realization of KBC involves two distinctive artificial musical "genres" meant to illustrate in sonic terms a similar opposition.

| Attribute | *Kiki* | *Bouba* |
|---|---|---|
| **Form** | Alternating accelerando rises and crazy section ("freak out") | Steady chorale |
| **Rhythm** | Accelerando and complex "free" rhythm, fast | Limited set of rhythmic durations, slow |
| **Pitch** | Modulo octave tuning system | Recursive subdivision tuning system |
| **Dynamics** | Fade ins and outs during accelerando and close of "freak out" sections | Single dynamic |
| **Voicing** | All voices in most of the time | Arch form envelope of voice density, starting and ending with single voice |
| **Timbre** | Percussive sounds alongside fast attack and decay bright pitched sounds. Second rise has an additional siren sound. | Slow attack and decay sounds with initial portamento and vibrato, with an accompanying dull thud |
| **Harmony** | Accidental coincidences only, no overall precepts | System of tonality, with a harmonic sequence built from relatively few possible chords |
| **Texture** | More homophonic in accelerando, heterogenous with independent voices in "freak out" sections | Homophonic, homogenous |
| **Expression** | Ensemble timing loose on accelerando, independent during "freak out" sections | Details of vibrato, portamento and "nervousness" (chance of sounding on a given chord) differ for each voice in the texture |
| **Space** | Little or no reverb | Very reverberant |

**Table 1**. Musical attributes of our realization of *Kiki* and *Bouba*.

for toolboxes of algorithms in an Aristotelean world, we pose the following tasks.

### 3.2 The discrimination task (unsupervised learning)

Given an unlabelled collection of music recordings from the music universe, build a system that determines there exist two categories in this music universe, *and* high-level (content) criteria that discriminate them. In machine learning, this can be seen as unsupervised learning, but ensuring discrimination is caused by content and not criteria that are irrelevant to the task.

### 3.3 The identification task (supervised learning)

Given a labelled collection of music recordings from the music universe, build a system that can learn to identify, using high-level (content) criteria, recordings of music (either from this music universe or from others) as being from *Kiki*, *Bouba*, or from neither. In machine learning, this can be seen as supervised learning, but ensuring identification is caused by content and not criteria that are irrelevant to the task.

### 3.4 The recognition task (retrieval)

Given a labelled collection of music recordings from this music universe, build a system that can recognize content in *real world music recordings* as being similar to contents in music from *Kiki*, *Bouba*, both, or neither. In information retrieval, this can be seen as relevance ranking.

### 3.5 The composition task (generation)

Given a labelled collection of music recordings from this music universe, build a system that composes music having content similar to music from *Kiki*, and/or music from *Bouba*. The rules that the system uses to create the music must themselves be meaningful. For example, a music analyst would find the program that generates the music to provide a high-level breakdown of the characteristics of a category. In one sense, this challenge is a necessary precursor to those above, in that a human composer must design the ground truth of the music universe. The production of a dataset of music recordings with algorithmic composition necessitates creation in real musical terms. The machine challenge here is to backwards engineer, or to learn in short, the compositional ability to work in the pre-established music universe. However, backwards engineering the compositional mechanisms of such a system, as an expert human musician can potentially do when encountering a musical style unfamiliar to them, is itself an important challenge of high-level musical understanding.

## 4. AN EXAMPLE REALIZATION OF KBC

We now present an example realization of KBC. We specify *Kiki* and *Bouba* via computer programs for algorithmic composition, which we use to create unlimited recordings of music from *Kiki* and *Bouba*, each varying subtly in the fine details (we discuss the practical range of this variation further below). Our computer program is written in the SuperCollider audio programming language [35], with SuperCollider used here in non-realtime mode for fast synthesis of music recordings (which in this case are monophonic digital audio files). We measure the speed of generation of music recordings to be around 60×real-time, so that one piece of around one minute can be created every second by our code. With this we easily created a multi-gigabyte dataset of ten hours, and could very easily create far more.

As *Kiki* and *Bouba* are designed here by humans, they are not independent of "real" music, even though they are fully specified via open source code. [3] Table 3 outlines properties of music from *Kiki* and *Bouba* with respect to some high and low level musical properties. This conveys a sense of why *Kiki* and *Bouba* are well-differentiated in musically meaningful ways. Figure 1 further attempts to illustrate the formal structure of the two styles, again as a demonstration of their distinctiveness. Although the musical description is not as simple as the visual manifestation of the original shapes of "kiki" and "bouba" [18,20], it was designed to avoid too much overlap of musical characteristics. Each output piece is around 40-60 seconds, since

[3] We make available this source code, as well as a few representative sound examples at the accompanying webpage: `http://composerprogrammer.com/kikibouba.html`.

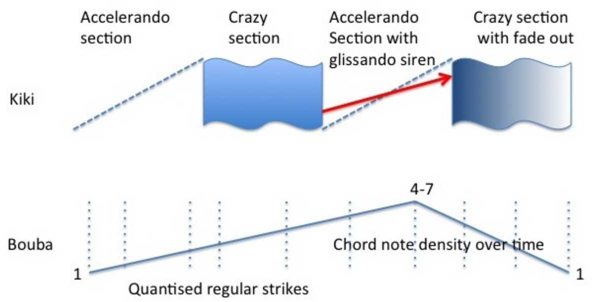


**Figure 1**. Comparative musical forms of our realization of music from *Kiki* and *Bouba* (labeled).

the actual length of sections is itself generative. It is beyond the scope of this article to discuss every detail of the code and the variability of output allowed, but this gives some idea. To anthropomorphise and allow a little literary conceit, our realization envisages music from *Kiki* to be ecstatic, chaotic and ritualistic, characterised by alternating build-ups (accelerando rises) and cathartic "freak-outs." Our realization envisages music from *Bouba* as an abstract choral funeral march, steady and affected.

### 4.1 An unacceptable solution

A typical approach to attempt to address an identification task is by computing a variety of low-level and short-time features from music recordings, modelling collections of these by probability distributions (bags of frames), and specifying criteria for classification, such as maximum likelihood. To this end, we use supervised learning to build a single nearest neighbor classifier trained with features computed from a dataset consisting of 250 recordings of music from *Kiki* and 250 from *Bouba*. As features, we first compute the number of zero crossings for 46.3 ms Hann-windowed audio frames, overlapped 50% across the entire recording. We then compute the mean and variance of the number of zero crossings from texture windows of 129 consecutive frames. Finally, we normalize the feature dimensions in the training dataset observations, and use the same normalization parameters to transform input observations. Figure 2 shows a scatter plot of these training dataset observations. To classify an input music recording as being of music from *Kiki* or *Bouba*, we use majority vote from the nearest neighbor classification of the first 10 consecutive texture windows.

We test the system using a stratified test dataset of 500 music recordings from *Kiki* or *Bouba*. For each input, we compare the system output to the ground truth. Our system produces a classification error of 0.00! It has thus successfully labeled all observations in the test dataset with the correct answer. However, this system is not a solution to the identification task of KBC, let alone the three other KBC tasks, *simply because it is not using high-level criteria (content)*. Of course, the statistics of low-level zero crossings across short-time frames has *something* to do with content [15], but this relationship is quite far removed and ambiguous. In other words, people listen to and describe music in terms related to key, tempo and timbre, but not zero crossings. Statistics of zero crossings are

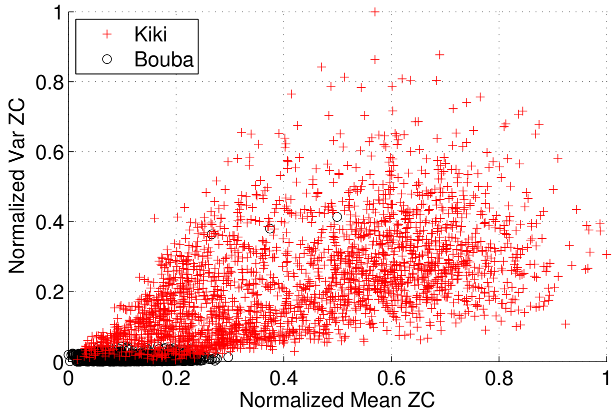


**Figure 2**. Scatter plot of features extracted from recordings of music from *Kiki* and *Bouba*.

not meaningful musical information for solving any task of KBC. That this feature contributes to the perfect figure of merit of this system, it does not illuminate what makes music *Kiki*, and what makes music *Bouba*.

### 4.2 An acceptable solution

As of the current time, we have yet to find any acceptable solution to our realization of KBC, or any of its tasks — which motivates this challenge. (Furthermore, as discussed below, the goal of KBC is not "a solution" but "solving.") We can, however, describe aspects of solutions acceptable for our specific realization of KBC. An acceptable solution to the discrimination task determines that in a set of music recordings from the music universe, there exist two different kinds of music, which are discriminable by high-level content, some of which are listed in Table 3, and shown in Fig. 1. An acceptable solution to the identification task determines for any given music recording whether its high-level contents are or are not consistent with all the musical attributes of *Kiki* or *Bouba*. An acceptable solution to the recognition task might recognize as *Bouba* characteristics the slow plodding rhythm, wailing timbre, and homophonic texture of some jazz funeral music. It might recognize as *Kiki* characteristics the glissando siren of some rave music, or the complex, unpredictable and ametrical rhythm of some free improvisation. It would recognize as not characteristic of either *Kiki* or *Bouba* the form of 12-bar blues. Finally, an acceptable solution to the composition task generates music that mimics particular characteristics of music from *Kiki* and *Bouba*.

## 5. DISCUSSION

In essence, KBC is a general exercise, of which we have provided one realization. KBC simplifies MGR — and music description in general — to the degree that many problems typically encountered in MIR research are not an issue, i.e., lack of data, copyright restrictions, cost and inaccuracy of ground truth, poor problem definition, and evaluations that lack validity with respect to meaningful musical understanding by machine. While most research in MGR searches for an Aristotelean categorization of real music (or the reverse engineering of the categorization used to create benchmark music datasets like GTZAN [30, 33]), it sustains most of the complexity inherent to the problem of MGR. KBC simplifies it to be Aristotelean and well-defined. Essentially, KBC defines categories of music as well-specified and open-source programs, which comports with an Aristotelean conception of music genre. This allows us to benefit from algorithmic composition since we can generate from these programs any quantity of data, free of copyright, and with a perfect ground truth and specified classification criteria.

It can be speculated that KBC is too much of a simplification of MGR, that defining music using programs has little "ecological validity," and thus that a solution to KBC will be of little use for music in the "real world." To the first claim, the tasks of KBC are much more complex than reproducing ground truth labels of datasets by any means — the implicit goal of the majority of work addressing MGR [27, 30] — *because solving the tasks requires machine listening*, i.e., "intelligent, automated processing of music" [8]. To the second claim, our realizations of music from *Kiki* and *Bouba* actually originate in higher-level musical processes defined by humans trained and practiced in music composition. Fundamentally, "algorithmic music" and "non-algorithmic music" is a false dichotomy; but this is not to say all algorithms create equally "valid" music. One non-sensical realization of KBC is defining music from *Kiki* and *Bouba* as 50 ms long compositions, each consisting of a single sine, but with frequencies separated by 1 Hz between the two categories. To the final claim, we emphasize an important distinction between "a solution to KBC" and "solving KBC." We are not claiming that, e.g., a system that has learned to discriminate between music from *Kiki* and *Bouba* will be useful for discriminating between "real" music using any two "real" genres. The system (the actual finished product and black box [29]) will likely be useless. Rather, *solving KBC* is the goal because this requires developing a system that demonstrates a capacity to listen to acoustic signals in ways that consider high level (musical) characteristics.

If one desires more complexity than KBC offers, one can conceive of a music universe with more than two categories, and/or various mixings of "base" categories, e.g., giving rise to cross-genres *Bouki* and *Kiba* (the code at our link already has the capacity to generate these hybrid forms). However, we contend the best strategy is to first demonstrably solve the simplest problems before tackling ones of increased difficulty. If the components of a proposed MGR system result in a system that does not solve KBC, then why should they be expected to result in a system that can discriminate between, or identify, or recognize, or compose music using "real" genres of music from a limited amount of data having a ground truth output by a complex culturally negotiated system that cannot be as unambiguously specified as *Kiki* and *Bouba*?

## 6. CONCLUSION

Simply described, content-based MIR research and development aims to design and deploy artificial systems that are useful for retrieving, using or making music content. The enormous number of published works [6, 14, 26, 38], not to mention the participation during the past ten years of MIREX, [4] show many researchers are striving to build machine listening systems that imitate the human ability to listen to, search for, and describe music. Examples of such research include music genre recognition, music mood recognition, music retrieval by similarity, cover song identification, and various aspects of music analysis, such as rhythmic and harmonic analysis, melody extraction, and segmentation. These pursuits, however, are hindered by several serious problems: a limited amount of data, the sharing of which is restricted by copyright; the problematic nature of obtaining "ground truth," and explicitly defining its relationship to music content; and a lack of validity in the evaluation of content-based MIR systems with respect to the task they are supposedly addressing. We are thus left to ask: *Have the simplest problems been demonstrably solved yet?*

In this paper, we show how algorithmic music composition facilitates limitless amounts of data, with perfect ground truth and no restricting copyright, thus holding appreciable potential for MIR research and development We propose the "Kiki-Bouba Challenge" (KBC) as a simplification of the problem of MGR, and produce an example realization of it facilitated by algorithmic composition. We do not present an acceptable solution to our realization of KBC, but discuss aspects of such a solution. We also illustrate an unacceptable solution, which fails to reveal anything relating to musical meaning even though it still perfectly labels a test dataset. We emphasize, *the goal of KBC is not the system itself, but in solving the challenge.* Solving KBC changes the incentive of research and development in content-based MIR from one of developing systems obtaining high figures of merit by any means, to one of developing systems obtaining high figures of merit by *relevant* means.

## 7. ACKNOWLEDGMENTS

We wish to acknowledge the anonymous reviewers who helped significantly in the revision of this paper. The work of BLS was supported in part by Independent Postdoc Grant 11-105218 from Det Frie Forskningsråd.

## 8. REFERENCES

[1] C. Ames. Automated composition in retrospect: 1956-1986. *Leonardo*, 20(2):169–185, 1987.

[2] C. Ames and M. Domino. Cybernetic Composer: an overview. In M. Balaban, K. Ebcioğlu, and O. Laske, editors, *Understanding Music with AI: Perspectives on Music Cognition*, pages 186–205. The AAAI Press/MIT Press, Menlo Park, CA, 1992.

[3] J.-J. Aucouturier and E. Bigand. Seven problems that keep MIR from attracting the interest of cognition and neuroscience. *J. Intell. Info. Systems*, 41(3):483–497, 2013.

[4] `http://www.music-ir.org/mirex`

[4] J.-J. Aucouturier and E. Pampalk. Introduction – from genres to tags: A little epistemology of music information retrieval research. *J. New Music Research*, 37(2):87–92, 2008.

[5] J. G. A. Barbedo and A. Lopes. Automatic genre classification of musical signals. *EURASIP J. Adv. Sig. Process.*, 2007:064960, 2007.

[6] T. Bertin-Mahieux, D. Eck, and M. Mandel. Automatic tagging of audio: The state-of-the-art. In W. Wang, editor, *Machine Audition: Principles, Algorithms and Systems*. IGI Publishing, 2010.

[7] G. C. Bowker and S. L. Star. *Sorting things out: Classification and its consequences*. The MIT Press, 1999.

[8] M. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. Content-based music information retrieval: Current directions and future challenges. *Proc. IEEE*, 96(4):668–696, Apr. 2008.

[9] D. Cope, editor. *Virtual Music : Computer Synthesis of Musical Style*. MIT Press, Cambridge, MA, 2001.

[10] D. C. Dennett. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. Simon and Schuster, 1996.

[11] S. Dubnov, G. Assayag, O. Lartillot, and G. Bejerano. Using machine learning methods for musical style modelling. *Computer*, pages 73–80, October 2003.

[12] F. Fabbri. A theory of musical genres: Two applications. In *Proc. Int. Conf. Popular Music Studies*, 1980.

[13] J. Frow. *Genre*. Routledge, New York, NY, USA, 2005.

[14] Z. Fu, G. Lu, K. M. Ting, and D. Zhang. A survey of audio-based music classification and annotation. *IEEE Trans. Multimedia*, 13(2):303–319, Apr. 2011.

[15] F. Gouyon, F. Pachet, and O. Delrue. On the use of zero-crossing rate for an application of classification of percussive sounds. In *Proc. DAFx*, 2000.

[16] Scott Johnson. The counterpoint of species. In J. Zorn, editor, *Arcana: Musicians on Music*, pages 18–58. Granary Books, Inc., New York, NY, 2000.

[17] M. Kassler. Towards musical information retrieval. *Perspectives of New Music*, 4(2):59–67, 1966.

[18] W. Köhler. *Gestalt Psychology*. Liveright, New York, 1929.

[19] J. McCormack and M. d'Inverno. *Computers and Creativity*. Springer-Verlag, Berlin Heidelberg, 2012.

[20] E. Milan, O. Iborra, M. J. de Cordoba, V. Juarez-Ramos, M. A. Rodríguez Artacho, and J. L. Rubio. The kiki-bouba effect a case of personification and ideaesthesia. *J. Consciousness Studies*, 20(1-2):1–2, 2013.

[21] E. R. Miranda, editor. *Readings in Music and Artificial Intelligence*. Harwood Academic Publishers, Amsterdam, 2000.

[22] G. Nierhaus. *Algorithmic Composition: Paradigms of Automated Music Generation*. Springer-Verlag/Wien, New York, NY, 2009.

[23] F. Pachet and D. Cazaly. A taxonomy of musical genres. In *Proc. Content-based Multimedia Information Access Conference*, Paris, France, Apr. 2000.

[24] M. Pearce, D. Meredith, and G. Wiggins. Motivations and methodologies for automation of the compositional process. *Musicae Scientiae*, 6(2):119–147, 2002.

[25] C. Roads. Research in music and artificial intelligence. *Computing Surveys*, 17(2), June 1985.

[26] B. L. Sturm. A survey of evaluation in music genre recognition. In *Proc. Adaptive Multimedia Retrieval*, Oct. 2012.

[27] B. L. Sturm. Classification accuracy is not enough: On the evaluation of music genre recognition systems. *J. Intell. Info. Systems*, 41(3):371–406, 2013.

[28] B. L. Sturm. Evaluating music emotion recognition: Lessons from music genre recognition? In *Proc. ICME*, 2013.

[29] B. L. Sturm. Making explicit the formalism underlying evaluation in music information retrieval research: A look at the MIREX automatic mood classification task. In *Post-proc. Computer Music Modeling and Research*, 2014.

[30] B. L. Sturm. The state of the art ten years after a state of the art: Future research in music information retrieval. *J. New Music Research*, 43(2):147–172, 2014.

[31] B. L. Sturm. A simple method to determine if a music information retrieval system is a "horse". *IEEE Trans. Multimedia*, 2014 (in press).

[32] B. L. Sturm, R. Bardeli, T. Langlois, and V. Emiya. Formalizing the problem of music description. In *Proc. ISMIR*, 2014.

[33] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Trans. Speech Audio Process.*, 10(5):293–302, July 2002.

[34] J. Urbano, M. Schedl, and X. Serra. Evaluation in music information retrieval. *J. Intell. Info. Systems*, 41(3):345–369, Dec. 2013.

[35] S. Wilson, D. Cottle, and N. Collins, editors. *The SuperCollider Book*. MIT Press, Cambridge, MA, 2011.

[36] I. Xenakis. *Formalized Music*. Pendragon Press, Stuyvesant, NY, 1992.

[37] Y.-H. Yang, D. Bogdanov, P. Herrera, and M. Sordo. Music retagging using label propagation and robust principal component analysis. In *Proc. Int. Conf. Companion on World Wide Web*, pages 869–876, 2012.

[38] Y.-H. Yang and H. H. Chen. *Music Emotion Recognition*. CRC Press, 2011.